

Automated Spatial Linkage

Vision and Requirements

Craig Wolff, M.S. Eng
CA Environmental Health Tracking Program
Environmental Health Investigations Branch
CA Department of Health Services
Impact Assessment, Inc.

EPHTN Context

- SND Emphasis: Secure network access to environmental public health data/metadata
- SND Tools discussion: help assess and analyze datasets individually
- California is examining EH-ified distributed services model
- What about linkage/integration methods as a service?
- What about environmental health linked data as an EPHTN data product?

What is EH Spatial Linkage?

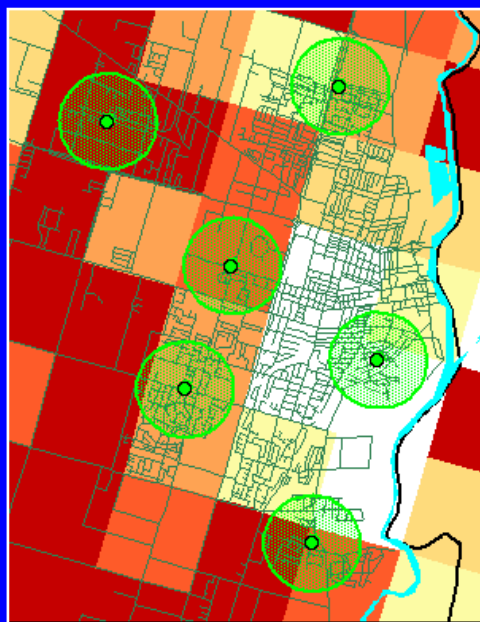
- The integration of environmental hazard and health data based on spatial relationships
- Spatial linkage begins after geocoding and ends before traditional statistical analysis of epidemiological associations
- Automated for remote procedure calls over the network



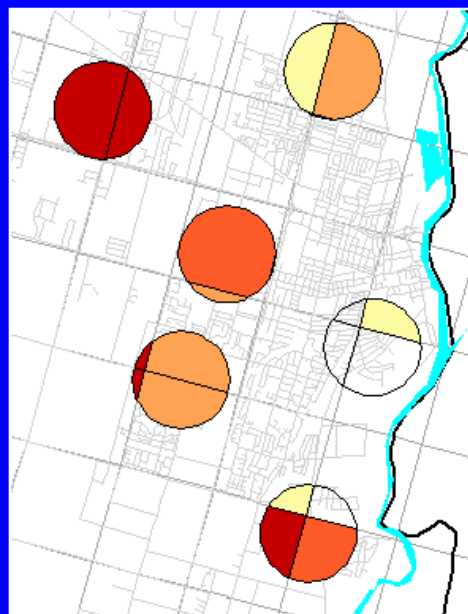
e.g. EH Spatial Linkage

A. Point events, B. Area events

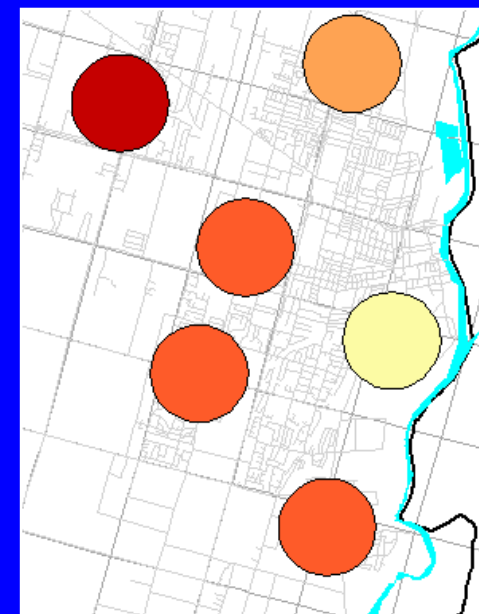
Buffer



Intersect



Summarize



Spatial linkage does not:

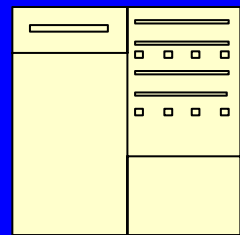
- Address issues of scientific validity or feasibility directly, though....
- Address data confidentiality and security issues directly, though....

General I/O Methodology

- EPHTN Gateway acts as intermediary for facilitating spatial linkage process & provides storage for linkage product
- Request
 - Health- or hazard-centric? 1° and 2° dataset?
 - Linkage method & parameters
 - Event definitions (spatial/temporal), 2° event metric
- Response
 - Operations performed, errors
 - Location of linkage product

e.g. Health-Centric Message Flow

Client



Step 6: Response indicating success/failure and where result table resides

Internet
(SSL)

Step 1: Request for spatial join – standard template

Environmental
2° Data Provider

spatial-enabled
hazard database

Step 5: Hazard
metrics returned
& inserted into
EH database

Step 4: Request
for spatial
transformation(s)
and hazard
metric
calculation

Health Department

Health 1°
Data Provider

spatial-enabled
health database

Step 2: Request for health events




Step 3: Health events returned

EPHTN
Gateway

spatial-enabled
EH database

Spatial Geometry/Dimension

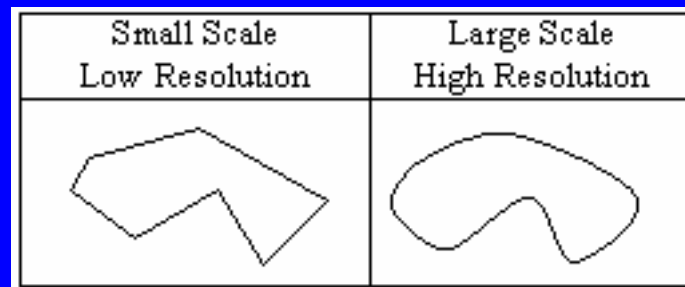
- Points, lines, areas, ~~raster?~~ ~~3-d?~~

Object Type	Points	Lines	Areas
Examples			
Health	Address <u>geocoded</u> asthma cases	Asthma cases along street segments	Asthma cases within census tracts
Hazard	Drinking water wells	Drinking water distribution pipes	Drinking water service areas

- Linkage methods should treat geometry generically

Spatial Resolution

- Data captured at large scale (high res) are more accurate than data captured at small scale (low res)



- Topological comparisons should be done on geometries having same or similar scale

Spatial Mismatch

- Environmental and health databases are not geo-referenced to the same spatial extent, geometry, AND scale
- For linkage, attributes of one or both datasets must be transformed to account for spatial mismatch
- System operators, not the system, must be thoughtful about quality/appropriateness

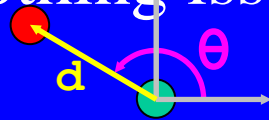
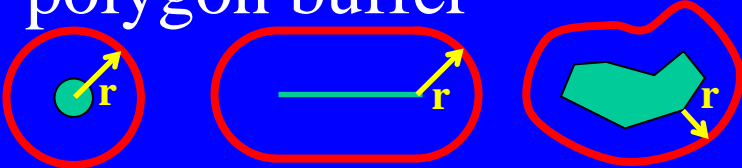
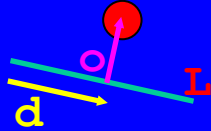
Event Definitions

- Event – spatio-temporal representation of health and environmental data to be linked
- 1° and 2° dataset – event is UID, geo-referenced geometry, and time component
- 2° dataset – event includes attribute(s) or metric(s) to be transformed to 1° geometry (AKA “secondary event metrics”)
- SND Content Modeling subgroup should hammer out EH event definitions as a standard

Transformation

- In EPHTN context, transformation is operation which alters spatial extent, geometry, or scale of health or hazard event
- Goal is to do 1+ (chain) transformations so that the secondary event metric can be merged with the primary event.
- Two types – parametric and topological

Parametric Transform

- Input geometry is transformed using one or more scalar input parameters
- Assist in scale mismatch and smoothing issues
- Point-to-point random shift
 
- Point/line/polygon-to-polygon buffer
 
- Line-to-point random distance “along” (with offset)
 
- Line-to-line or area-to-area generalization (node removal at regular intervals)

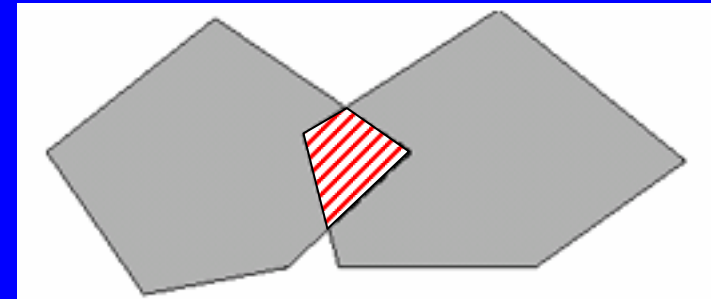
Topological Transform

- Input geometry is transformed by comparing it topologically to another input geometry
- Scale mismatch between input datasets should be minimized or non-existent
- Topological operators of interest: nearest and intersect

Intersection

- Dimensional Intersection Model: Interior, Boundary, and Exterior (9 combinations for two input geometries)*

		Area		
		Interior	Boundary	Exterior
Area	Interior	2		
	Boundary			
	Exterior			



- Keep it simple. Neglect unlikely or uninteresting combinations
- Named spatial relationships (relational operators) of interest: crosses, within, and overlaps. Neglect touches & disjoint?

Intersection Output Scalar

- Proportional or absolute value describing secondary input geometry's relationship (distance or area) to transformed output geometry \hat{Q}
- Used in merging 2° event metric with 1° event
- Example: $\text{Area}(1^\circ) = A_1$, $\text{Area}(2^\circ) = A_2$,
 $\text{Area}(1^\circ \cap 2^\circ) = A_{12} \rightarrow \hat{Q} = A_2 / A_{12}$

Merge the Metric

- Summary operator applied to corresponding secondary event metrics: sum, minimum, maximum, average, frequency, standard deviation
- Output scalar applied to secondary event metric before summarizing about primary event

Software Requirements

- Spatial linkage object model: Interfaces, objects, methods which generically treat health/hazard events and transformations
- I/O entry/exit points: software objects should be serializable to XML schema (SOAP? PHIN-MS?)
- Geographic feature data passed over the wire should follow OGC GML
- Toolkit to assist Tracking partners in exposing their data holdings as spatial linkage services

Next Steps

- Refine vision and requirements... Tools Subgroup (early 2005)
- Lessons to-be-learned from California pilot projects
 - Automated pesticide linkage (early 2005)
 - Automated air contaminant linkage (mid 2005)
 - Automated traffic hazard linkage (mid 2005)
- Contractor completes analysis and implements abstract API components (future projects)



Discussion